A Study of the Feasibility of Using Quantitative Comparison Type of Questions in Mathematics Examinations in Hong Kong

by

Wong Yuen Ting

School of Education

The Chinese University of Hong Kong

1982

ABSTRACT

This study is a feasibility study of using quantitative comparison (QC) type of questions in mathematics examinations in Hong Kong. QC items are objective items which involve the comparison of the magnitude of two quantities. Examinees are required to determine whether one of the quantity is greater than, equal to or less than the other quantity; or there is no quantitative relationship between the two quantities. This type of items first appeared in the mathematics part of the Scholastic Aptitude Test run by the Educational Testing Service in the United States.

The subjects were 138 F.5 students in a subsidised Anglo-Chinese co-educational school. The subjects were divided into two matched groups according to their sexes, language and mathematics abilities. Subjects in one group took a test which was made up of traditional multiple-choice (MC) items whereas subjects in the other group took another test which contained matched items in the QC format. No significant difference was found between the two tests in the mean score, the discriminating power, the reliability and the rate of working on MC and QC items. The concurrent validity for the two tests were found to be significant different.

It was found that QC items offer a reliable and valid test of mathematics achievement at the Certificate of Education level. However, it must be noted that QC items should be used together with MC items and they cannot replace MC items. It was also suggested that, before introducing QC items in mathematics examinations in Hong Kong, more research studies should be done in relating to the effect of the use of calculator, mathematics topics and the examinees' abilities on QC items.

ACKNOWLEDGEMENTS

I wish to acknowledge the following people and institutions for the help they offered over the various aspects in this study :

(1) Dr S.C. Cheng, my supervisor, for his guidance and advice,

(2) Hong Kong Examinations Authority, for the permission to use mathematics items from past Hong Kong Certification of Education Examination question papers,

(3) Mr C.C. Choi, Senior Examination Officer of the Hong Kong Examinations Authority, for his assistance in the analysis of data,

(4) Mr S.K. San, the principal of Chinese Y.M.C.A. Secondary School, for his arrangements in connection with the administration of the tests and his permission in allowing his teachers and students to take part in the study,

(5) Mr K.K. Mok, Mr Y.L. Ho and Mr K.Y. Pun, mathematics teachers in Chinese Y.M.C.A. Secondary School, for their assistance in test invigilation,

(6) all the F.5 students in the Chinese Y.M.C.A. Secondary School, for their kind and patient cooperation in taking the tests,

(7) Mr H.C. Law, my colleague at the Hong Kong Examinations Authority, for his advice on the design of the experiment,

(8) Mrs Camilla Wong, for all the typing work.

CONTENTS

		Page
I	Background of the Study	1
	A. Objective Test	1
	B. Quantitative Comparison Type of Items	4
	1. A new Type of Objective Items	4
	2. Advantages and Limitations of QC items	5
II	Statement of the Problem	8
	A. Multiple-Choice Testing in Hong Kong	8
	B. Use of Quantitative Comparison Items in Hong Kong	11
	C. Hypotheses	12
III	Literature Review	13
	1. Effect of Item Types	13
	2. Effect of Language Abilities	16
	3. Effect of the Number of Options	17
IV	Design of the Experiment	19
	A. The Experiment	19
	B. Testing Instrument	20
	C. Administration of the Tests	21
	D. Analysis of Data	23

V Results and Discussion

A. Findings 27

1. Mean Score 27

2. Discriminating Power 27

3. Reliability 28

4. Validity 28

5. Time Taken to Complete Equal Number of MC and QC Items 28

B. Discussion 29

1. Mean Score 29

2. Discriminating Power 30

3. Reliability 31

4. Validity 31

5. Time Taken to Complete Equal Number of MC and QC Items 32

VI Limitations of the Experiment 34

VII Conclusion and Suggestions 36

Bibliography 38

Appendices (I to IV)

Chapter I

BACKGROUND OF THE STUDY

A. Objective Test

All objective tests have a common characteristic of making up of items which have predetermined keys. The key to each item is established when the item writer writes the item. This means that scoring an objective test is very mechanical and can be done by a person who knows nothing about the subject matter of the test or even by test-scoring machines. The scoring will also be independent of the scorer. Whoever scores the test, the score will be the same if the test is scored according to the predetermined keys. This uniformity in scoring as well as the speed in scoring by test-scoring machines attribute to the success and the popularity of objective tests.

There are different types of objective items. Each type is characterised by : (i) the arrangement of words, phrases, sentences or symbols composing the items; (ii) the directions given to the examinees for response to the items; and (iii) the provision made for recording the response. Among the wide variety of objective items, the Multiple-Choice (MC) type is the most popular. An MC item consists of an item stem and two or more responses. The stem is an essential part of an item. It presents the problem and asks for the solution to the problem. It is either in the form of an introductory question or an incomplete statement. For some items, information relating to the problems presented is also included in the stems. The responses are the suggested answers to the question or the suggested completions to the statement. Among the responses, the correct one is called the key and the incorrect ones are called the distractors. The distractors together with the key are called the options. In an MC item, an examinee is required to choose an option as his answer.

Thorndike and Hagen (1961) referred to MC items as the most effective of the objective item types. It is effective for measuring information, vocabulary, comprehension, application of principles, or ability to interpret data. In fact, it can be used to test practically any educational objective that can be measured by a pencil-and-paper test except the ability to organize and present materials. Ebel (1979) referred to MC items as the most highly regarded and widely used form of objective test item. MC items can be used to measure important educational objectives. According to Wesman (1971), MC items are the most popular type of objective items in current use. They are free from many of the weaknesses inherent in other types. They are adaptable to a wide variety of topics.

While they have often been used to measure superficial verbal associations and insignificant factual details, they can also be used with great skill and effectiveness to measure complex abilities and fundamental understandings. In fact, Anderson (1979) has shown how different types of MC items were used in the assessment of an analysis course in a first-year university mathematics curriculum in the United Kingdom.

However, MC items have their limitations. One of the difficulties that all item writers have encountered in writing MC items is the finding of suitable and good options. Often, items with very good stems have to be discarded just because of the failure in finding a sufficient number of good distractors. Moreover, from the writer's own experience, much of the time taken in writing an item is spent in constructing the distractors. A common criticism on MC items is that an examinee may obtain the answer by substituting each of the options in the problem. Thus an examinee may answer an item correctly even though he may not know how to solve the problem. It is desirable to have objective items which can eliminate or at least minimise the above limitations.

B. Quantitative Comparison Type of Items

1. A New Type of Objective Items

According to Braswell (1978), the Educational Testing Service (ETS) in the United States introduced a new type of objective test items in the mathematics part of its Scholastic Aptitude Test (SAT) in 1975. This new type of items were called the quantitative comparison (QC) items. QC items were used to replace items called the data sufficiency (DS) type, which were used in the SAT from 1959 to 1974. After 1974, all the items of the mathematics part in the SAT are either of the QC type or ordinary MC type.

In an QC item, examinees are presented with two quantities, one in column A and the other in column B. Examinees are then required to compare the magnitude of these two quantities and are instructed to mark A, B and C as the answer if the quantity in column A is respectively greater than, less than or equal to the quantity in column B. If there is not enough information to determine a quantitative relationship between the two quantities, then D should be marked as the answer. In some QC items, information in the form of sentences and/or figures concerning one or both quantities may also be provided. Though the number of possible answers of QC items is four whereas ordinary MC items have five options, the answering format are similar for the two types. Therefore, answer sheets for ordinary MC items can also be used for QC items.

2. Advantages and Limitations of QC Items

In QC items, the trouble of finding good and suitable distractors would be much reduced as there is no distractor and item writers only have to find two quantities for comparison. Moreover, one of the quantities is usually the answer to the problem presented. These two factors make it easier to write an QC item than an MC item.

Braswell (1978), in an article on the mathematics portion of the SAT, asserted that QC items required less time to answer, involved less reading, and required less computation than ordinary MC items. However, he quoted no evidence to support his assertions.

In 1972, Pike and Evans of the ETS conducted research studies on QC items. They found out that QC items contributed to the predictive validity of the SAT. This increase in predictive validity was shown by an increase in correlation of the test scores of the SAT with college grades. Moreover, individual QC items were generally less dependent on verbal skills than were DS items. As a result, the correlation between the verbal and mathematical scores was reduced. This was desirable as the mathematical score represented a more genuine measure of an examinee's mathematical ability.

One of the criticisms on MC items is that an examinee may obtain the answer by substituting each of the options in the problem. This substitution method is especially common in mathematics. Special care has to be taken in constructing an item in order to prevent examinees from getting the answers by this method. For QC items, the problem of getting the answer by substitution does not exist since the examinees are only required to compare the given quantities.

In QC items, the two quantities presented must be quantitative in order that a quantitative comparison can be made. This restriction makes it impossible to test some mathematics topics using QC items. These include factorization and simplification of algebraic expressions, trigonometric identities, algebraic inequalities involving two or more unknowns. Those problems involving graphs and properties of numbers also cannot be tested using QC items.

Some other problems which have the answers expressed in terms of unknown quantities represented by letters may be tested by QC items by substituting numerals for the unknown quantities. However, in such a change, the difficulty of an item would be lowered. Thus QC items cannot replace traditional MC items and must be used together with other types of MC items in an objective test.

In the HKCEE, candidates are allowed to use electronic calculators in all the examination papers. Since QC item involves the comparison of two quantities, the use of calculators may change the solution of a mathematical problem into mere button-pushing. Either examinees should not be allowed to use calculators or special care has to be taken not to include those problems which depend very much on calculator operations.

Besides the coverage of the syllabus and the use of calculator problems, QC items also have the disadvantage of being susceptible to relatively short-term instruction (Pike and Evans, 1973). It was found that because of the relatively more complex format of QC items, they were more susceptible to special coaching than ordinary MC items.

Chapter II

STATEMENT OF THE PROBLEM

A. Multiple-Choice Testing in Hong Kong

Most of the major* subjects in the Hong Kong Certificate of Education Examination (HKCEE) are examined in two papers. One of the papers makes use of the essay-type questions (commonly referred to as the conventional paper). In this paper, examinees have to organize their own answers and present the answers in their own words. The answers are then marked by a panel of markers according to a marking scheme. However, it must be pointed out that the answers given by an examinee are marked by only one marker. Although regular checkmarking is used to monitor the standard of marking, and statistical methods are available for adjusting the marks awarded by individual markers, individual differences among markers might affect the marking of a conventional paper to such an extent that inconsistency may still occur among markers. Sometimes, this happens even for the same person marking at different times. The other paper consists of MC items (usually referred to as the MC paper). In this paper, examinees mark their answers on specially prepared answer sheets and their answers are scored by electronic scoring machines.

* Major Subjects refer to subjects in which the number of candidates taking the subjects is large (usually above 10 000).

Though MC items are now widely used in Hong Kong in public examinations as well as internal assessments in schools, the history of MC testing in Hong Kong is not very long. According to Lee (1980), MC items were first used in public examinations in Hong Kong in 1963 in the subject of English Language in the former Hong Kong English School Certificate Examination. However, it was not until 1969 that four other subjects (Biology, Chemistry, Chinese History and Mathematics) also included MC items in their examinations. In 1974, when the Hong Kong Certificate of Education (English) Examination and the Hong Kong Certificate of Education (Chinese) Examination were amalgamated, Chinese versions of MC papers were first found in the subjects Economics and Public Affairs, History and Physics. Since then, MC items have become more and more popular in Hong Kong, and in 1981 a total of 31 subjects (counting alternative syllabuses, English and Chinese versions separately) in the HKCEE were examined with an MC component.

Mathematics was one of the first five subjects which included MC items in the Hong Kong English School Certificate Examination in 1969. The correlation between the marks of the conventional paper and the MC paper has been very high. Lee (1980) pointed out that the correlation between the two papers in Mathematics in the 1976 HKCEE was so high (0.914) that it was wondered whether the MC paper alone would be sufficient for testing and differentiating the candidates, thus saving thousands of man-hours of manual marking. Besides the high correlation with the conventional paper, the importance of the MC paper in Mathematics can also be seen from the fact that the MC paper marks make up 40% of a candidate's total subject marks.

There are various types of MC items. Some types can only be used in some subjects. In mathematics, most of the items contain sentences and/or figures which provide information relating to the problem presented. A review of the two books of sample MC papers published respectively by the Hong Kong Certificate of Education Board in 1976 and the Hong Kong Examinations Authority in 1980 showed four types of MC items used in mathematics: (a) Direct Question Type, (b) Incomplete Statement Type, (c) Negative Type and (d) Multiple Completion Type.

In all the MC items in mathematics, the number of options is five. It should also be noted that most of the items in the HKCEE are of the direct question type and the incomplete statement type. Items of the negative type and the multiple completion type only make up a very small fraction in an MC paper.

B. Use of Quantitative Comparison Items in Hong Kong

It has been almost 20 years since MC items were first used in public examinations in Hong Kong. People now accept that MC items offer a reliable and a valid measurement of a candidate's ability in a subject. This is especially true for mathematics and science subjects as the correlation between the scores of the MC paper and conventional paper in mathematics and sciences in the HKCEE has been very high. Since QC items have many advantages over MC items and yet retain the characteristics of objective items, one may wonder whether it would be possible to use QC items in local secondary public examination. Though several researches had been done in USA and QC items have been included in the mathematics part of the SAT run by ETS for a number of years, very few teachers and students here in Hong Kong have even heard of this type of items. It is the purpose of this study to examine the feasibility of using QC items in local secondary public examination.

Since the effectiveness of objective items are characterised mainly by the difficulty, discrimination, reliability and validity of the items, the present study will be concerned with the comparison of MC and QC items in these aspects. Moreover, the time taken to complete an item would affect the number of items to be included in an examination paper. Thus the average times taken to complete an QC item and an MC item would be compared.

C. <u>Hypotheses</u>

For QC and MC items of the same content,

(1) there is no significant difference between the mean scores of QC and MC items;

(2) there is no significant difference between the discriminating power of QC and MC items;

(3) there is no significant difference between the reliability of QC and MC items;

(4) there is no significant difference between the validity of QC and MC items;

(5) there is no significant difference between the times taken to attempt an equal number of QC and MC items.

Chapter III

LITERATURE REVIEW

Before the results of an examination paper are used to judge the performance of the examinees, one need to know how much reliance can be place on the results obtained. There are two important factors which attribute to a good test. These two factors are the reliability and the validity of a test and they are closely related to each other. A reliable test is one that yields consistent results whereas a valid test is one that demonstrably measures what it was intended to measure. The factors affecting the reliability of a test include the length of the test, homogeneity of content, discrimination, difficulty, ability of examinees and time limit. However, validity of a test is not an intrinsic property of the test but depends on the examinees and the purpose of the test; and the judgement of the validity of a test is subjective and personal.

Numerous researches have been done to study the effect of various factors affecting the effectiveness of a test. Some of these factors are discussed in greater details below.

1. Effect of Item Types

According to Tinkleman (1971), "It is possible to measure the same test objective with a variety of different item types." However, different test authors have different convictions. For example, in a mathematics test, some test constructors are concerned with the possibility that the examinees may be able to get the correct answer by substituting each of the options in turn; hence, they include in every question an option to the effect that the correct answer is none of those listed. Others have a strong objection towards none-of-the-above type of option. Some others insist that all the items must be of the completion type in which no option is given and the examinees have to work out their own answers.

Wesman (1971), in an article titled "Writing the Test Item", quoted some researches on the effect of item types. Among the researchers quoted were Frederiksen and Satter (1953), Rimland and Zwerski (1962), Wesman and Bennet (1946), Boynton (1950). Frederiksen and Satter (1953) compared the difficulty indices of arithmetic computation problems presented in MC and completion types of items. They found that the difficulty of the problems was very similar under the two conditions. Rimland and Zwerski (1962) compared the frequencies of choosing distractors in MC and completion items in arithmetic reasoning and computation problems. Again, similar frequencies were found under the two conditions. Wesman and Bennet (1946) investigated the use of none-of-these as an option in a vocabulary test as well as an arithmetic test. Two tests, one containing ordinary 5-option items and the other containing (among the five options) a none-of-these option, were administered to two matched groups. No difference in difficulty or in correlation with an external criterion was found. However, Boynton (1950) found that items with the none-of-these option in a spelling test were more difficult than items without it.

Benson & Crocker (1979) found that the performance of candidates in an objective test was influenced significantly by the item type and the reading ability of the candidates. The items in a health science test were written in true-false (TF), MC and matching formats. It was found that students performed best in the matching type and worst in the TF type. Another result worth noting is that students with higher reading ability performed better than students belonging to the lower reading ability group.

A special type of MC item has been constructed to assess mainly the process skill in problem solving in mathematics. Instead of numerical answers, the options are several possible procedures for solving the problems. Forsyth and Spratt (1980) investigated the effect of this new type of MC items. It was found that this new item type was more difficult, had a lower discrimination and a lower reliability value than ordinary MC items. As for the validity, there was no definite answer whether the new and the ordinary MC items were measuring the same thing.

Frisbie (1973) conducted a study to compare the reliability and concurrent validity of both MC and TF items. It was found that the TF items were significantly less reliable than the MC items but these two types of items did tend to measure the same thing.

The effects of special instruction course are also affected by the item types. Evan and Pike (1973) compared the effects of instruction for three mathematics item formats, namely QC, DS and regular multiple-choice (RM) types. It was found that QC items were most affected by special coaching, followed by the DS and the RM type respectively.

2. Effect of Language Abilities

It is generally recognized that linguistic abilities affect performance in mathematics. Aiken (1972), in a review article, quoted some researches in which performances in mathematics were found to be positively correlated with general language abilities. These included the researches by Balow (1964), Chase (1960), Cleland and Toussaint (1962), Cottrell (1968), Erickson (1958), Ivanoff, DeWane and Praem (1965), Muscio (1962), and Pitts (1952). Most of these studies involved children in the intermediate grades and correlations between general reading ability and mathematics achievement were computed. Correlation coefficients obtained ranged from 0.40 to 0.86. However, when Johnston (1949) and Henney (1969) conducted studies concerning specific reading abilities in mathematics, their results did not demonstrate superior predictive validity for measures of specific reading abilities when compared with measures of general reading abilities.

Braswell (1978) of the ETS reported that QC items were individually less dependent on verbal skill than the DS items. After the inclusion of QC items in the SAT, it was found that the correlation between the verbal and the mathematical scores had been reduced.

3. Effect of the Number of Options

According to Wesman (1971), there was no definite answer as to the optimum number of distractors in an MC item although theoretically speaking, adding the number of distractors to an item would increase its reliability. The practical difficulty of finding good distractors would result in a futile attempt to include a large number of distractors. Moreover, more distractors would mean more time for reading, thus fewer items may be included in the test. In practice, most item writers write four- or five-option items.

Ramos and Stern (1973) compared the difference between four-option items with five-option items rescored as four-option items after removing the least popular distractor. It was found that four-option items were significantly less reliable than five-option items.

Lee (1980) showed that for four shortened tests (30 minutes each) of MC items in English Language, Chinese Language, Chemistry and Geography, the removal of the least popular option in 5-option MC items has no significant effect on the test results, the discrimination, the reliability and the validity.

Chapter IV

DESIGN OF THE EXPERIMENT

(A) The Experiment

The subjects in the study were 4 classes of F.5 students in a subsidised Anglo-Chinese co-educational school. There were altogether 144 subjects and they were divided into two groups A and B. Each group took a different test and their performance in the tests were compared. As the performance of the examinees might be affected by the sex, language ability and mathematics ability of the examinees, the subjects in the two groups were matched according to their sexes, language and mathematics abilities. The language and mathematics abilities were respectively measured by the subjects' achievement in English Language and Mathematics in their F.5 half-yearly examination. As all the items in the two tests were written in English, English Language was chosen as a measure of the subjects' language abilities. Although the same question papers were used for the four classes in the internal assessment exercise, there might be differences between classes as the four classes were taught by different teachers and thus the conventional papers used in the internal assessment exercise were marked by different persons. In order to cater for the differences between classes, matching of subjects was done within a class. For example, if a male subject was placed in Group A, than another male subject from the same class with comparable abilities in English and Mathematics would be placed in Group B.

B. Testing Instrument

There were two objective tests, Test A and Test B. Subjects in Group A would take Test A whereas subjects in Group B would take Test B. Each test contained 36 items and was of duration 1-hour. Therefore, the average time allocated for one item was $1\frac{2}{3}$ minutes and was approximately the same as that in the HKCEE.

All the items in Test A were MC items and were taken from past HKCEE papers. Test B contained QC items which were matched with the MC items in Test A. The order of appearance of the items were the same in the two tests. Thus the n th item in Test A presented the same problem as the n th item in Test B. The only difference was that the problems were presented by MC items in Test A whereas in Test B they were presented by QC items.

In the HKCEE, all the mathematics items are of the 5-option type. However, in an QC item there are only four possible answers. In order to compare the performance on QC and MC items, the two types of items should have the same number of possible answers. All the items in Test A, after selected from past HKCEE papers, had the least popular distractor deleted. Thus, each item contained only 4 options. In selecting the MC items from past HKCEE papers to form Test A, the facility, the discrimination index and syllabus coverage of the items had been taken into account. Although it would be impossible to cover the whole mathematics syllabus in the HKCEE using QC items, special effort had been made to cover as many topics as possible.

C. Administration of the Tests

The two tests were administered to the four classes of subjects in their own classrooms at the same time by the writer and three mathematics teachers from the school. Before administering the tests, the three teachers were briefed by the writer on the procedure of administering the tests. The nature of QC items were explained to the teachers in great detail with the help of examples. To facilitate the administration of the tests, an instruction sheet for those invigilating the tests had been prepared and distributed to the three teachers. The instruction sheet were to ensure uniform administrative procedures were observed in the four different classrooms.

The two tests were printed on papers in different colours with answer sheets of matching colours. Question papers and answer sheets for Test A were printed in white while those for Test B were in green. For the convenience of the invigilators and the subjects, the name of each subject had been written by the writer on the answer sheets before the test.

Before distributing the test papers and answer sheets, each invigilator gave a short introduction on the test. The subjects were told that the tests they were going to take concerned a research study and their results in the tests would not affect their results in their internal assessment and external examination. Moreover, the result of individual subjects would be treated in confidence and would not be disclosed to other persons. However, the subjects were encouraged to do their best in the test.

In distributing the test papers and answer sheets, invigilators called out the names of the subjects and each subject was given a test paper and an answer sheet. The subjects were reminded to check whether they received the right answer sheets and whether the answer sheets and the question papers were of the same colour. They were then told to read the instructions to candidates on the cover of the question papers. Details regarding the jotting down of the number of items attempted at the end of the first 15 minutes, 30 minutes and 45 minutes after start were given. The nature of QC items was then explained to all the subjects in Group B by referring to the examples on the front page of Test B.

After making sure that every subject had no query about the test, the subjects were told to start working. At the end of the first 15 minutes, 30 minutes and 45 minutes after start, subjects were told to put down the number of items attempted in the appropriate boxes on the answer sheets. Five minutes before the end of the test, subjects were reminded of the time left. After the test, answer sheets and question papers were collected separately and subjects were thanked for their co-operation and assistance in this research study.

D. Analysis of Data

Since all the MC items were taken from past HKCEE papers and in selecting the items, the facility (mean = 0.600) and discrimination index (mean biserial coefficient of correlation = 0.635) of the items had been taken into account, the MC items were treated as the control items. Subjects in Group A, taking Test A, belonged to the control group whereas subjects in Group B belonged to the experimental group. The results of the control group and the experimental group were compared in the following aspects in order to test the five hypotheses listed in Chapter II:

(1) Mean Score

The mean score in each test was measured by the mean number of correct items in the test. The mean scores in the two tests were then compared using t-test.

(2) Discrimination Power

The discrimination power of the tests was measured by the biserial coefficient of correlation, r_b . In order to compare the discrimination of the two tests, Fisher's Z-transformation procedure was applied to r_b. The transformation equation is

$$Z = \frac{1}{2} \log_e \left[\frac{1 + r_b}{1 - r_b} \right]$$

$$\text{or} \quad Z = \tanh^{-1} r_b \ .$$

The sampling distribution of Fisher's Z is normal, therefore the sampling distribution of $Z_1 - Z_2$ (Z_1 and Z_2 are the Fisher's Z coefficients for Test A and Test B respectively) is also normal. The discrimination of the two tests was compared by computing

$$\frac{Z_1 - Z_2}{SE},$$

where $SE = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$.

= Standard error of a difference between two independent Z coefficients from two samples consisting of N_1 and N_2 subjects respectively.

(3) Reliability

The reliability of the tests was measured by the Kuder-Richardson Formula 20 (KR-20). The KR-20 coefficients for the two tests were then compared using a procedure developed by Feldt (1969).

(4) Validity

In this study validity referred to the concurrent validity and was measured by the correlations between the scores in each of the two tests with the subjects' internal assessment results in mathematics. The correlations for the two tests were compared using the same procedure as that used for comparing the discrimination power.

(5) Time Required

The times taken to attempt an equal number of QC and MC items were measured by the mean number of items attempted at 15 minutes, 30 minutes and 45 minutes after start. The mean numbers of QC and MC items attempted were compared using t-test.

Chapter V

RESULTS AND DISCUSSION

Of the 144 subjects in the four classes, two were absent and two were late on the day of the experiment. As the subjects in the two groups were matched, two of the subjects matched with the absentees and the late-comers had to be excluded from the analysis. Finally, the results of 138 subjects, with 69 in each group, were analysed.

(A) Findings

1. Mean Score

Table 1

Means and Standard Deviations of the Number of Correct Items in the Two Tests.

	Control Group (MC)	Experimental Group (QC)
Mean	23.67 (65.75%)	21.77 (60.47%)
S.D.	6.34	6.54

2. Discriminating Power

Table 2

Means and Standard Deviations of the Biserial Coefficient of Correlation (r_b) of the Two Tests.

	Control Group (MC)	Experimental Group (QC)
Mean	0.5400	0.5274
S.D.	0.1569	0.1643

3. Reliability

KR-20 coefficients for the Two Tests:

Control Group (MC) = 0.846

Experimental Group (QC) = 0.850.

4. Validity

Corelation Coefficients between Internal Assessment Results and Test Scores:

Control Group (MC) = 0.888

Experimental Group (QC) = 0.736.

5. Time Taken to Complete Equal Number of MC and QC Items

Table 3

Numbers of Items Attempted at Various Times after start.

		Time in Minutes		
Test Form		15	30	45
Test A (MC)	Mean	10.67	23.77	32.86
	S.D.	2.93	6.60	3.85
Test B (QC)	Mean	11.42	25.72	34.65
	S.D.	2.99	5.32	2.36

B. Discussion

1. Mean Score

Table 1 shows the means and standard deviations of the number of correct items in Test A (MC) and Test B (QC). No significant difference ($p > 0.05$) was found between the means of the number of correct items in the two tests. However, a closer examination of the item analysis of the two tests revealed that out of the 36 items in Test A (MC), seven items had p-values (percentage of subjects answering an item correctly) greater than the p-values of matched items in Test B (QC) by 0.2 or higher. There was only one item in Test B (QC) which had a p-value greater than the p-value of the corresponding item in Test A (MC) by 0.2 or higher. Although the difference between the scores in the two tests was not statistically significant, some individual items (about 20%) were more difficult when presented in the QC format than in the MC format. The lower p-values of some QC items might be caused by the subjects' lower ability in comparing quantities involving fractions or negative numbers. Moreover, the difficulty of an QC item also depends on how the item was modified from an MC item. For example in Item 13, it would be easier to get the answer, without any computation, in Test B (QC) than in Test A (MC). The situation was reversed in Item 19. Thus changing an MC item into an QC item might increase or decrease its difficulty depending on how the item is modified. There does not seem to be a definite answer to the question whether MC or QC items would be easier.

2. Discriminating Power

Some measurement experts recommend that the mean p-value for an MC test should be about 0.60 in order to achieve maximum discrimination (see, for example Henrysson (1971)). The mean p-values for the two tests in the present study were quite close to this standard and the influence on the discrimination power by the difficulty of the items would be quite small. A comparison of the Fisher's Z coefficients of the two biserial coefficients of correlation of the two tests showed no significant difference. However, there were 6 items in Test A (MC) which showed a much greater r_b than matched items in Test B (QC). For Test B (QC), there were also 6 items which had a much greater r_b. The difference between the discrimination power of some MC and QC items might lead one to infer that some problems are more suitable to be examined in QC format as these problems would be more discriminating when presented in the QC format than the MC format. Of course, there are problems which are better examined in MC format, and there were about 10 items in the tests in which the discriminating power of QC and MC formats were very similar.

3. Reliability

The Kuder-Richardson 20(KR-20) coefficients for Test A (MC) and Test B (QC) were 0.846 and 0.850 respectively. Comparison using a procedure developed by Feldt (1969) showed no significant difference between the two KR-20 coefficients. Since all the MC items in Test A (MC) were taken from past HKCEE papers, it can be inferred that QC items are reliable in the testing of the mathematics achievement for subjects at this level.

4. Validity

In the present study, validity refers to concurrent validity and was measured by the correlation coefficient between an subject's mean score in a test with his/her internal assessment result. The correlation coefficients for Test A (MC) and Test B (QC) were respectively 0.888 and 0.736. The difference between these two values were found to be statistically significant ($p < 1\%$). The internal assessment was made up of two papers, a conventional paper and an MC paper. The MC paper in the internal assessment was of the same format as that in the HKCEE and was thus similar to Test A. This similarity between Test A (MC) and part of the internal assessment may account for the higher ~~reliability~~ correlation coefficient obtained for Test A (MC). However, the difference in validity for the two tests cannot justify the assertion that the two types of items were measuring different things. An QC item present the same mathematics problem as the matched MC item. The only difference is that in an QC item, the answer can only be determined after the two quantities have been compared. Thus, subjects taking Test B (QC) had to take one more step to get the answers. This extra step required a different ability from the subjects, namely, the ability to perform a quantitative comparison between two quantities. This accounted for the difference in validity for the two tests.

5. Time Taken to Complete Equal Number of MC and QC Items

The number of items attempted at three specified times after start was used as a measure of the average time taken to complete equal number of MC and QC items. The numbers of items attempted at 15 minutes and 30 minutes after start showed no significant difference for the two tests ($p > 0.05$). However, the number of items attempted at 45 minutes after start were found to be statistically different ($p < 0.01$). This was caused by the smaller standard deviations (compared with the means). The small standard deviations were due to the fact that many subjects had completed the whole test in less than 45 minutes and they were instructed to put down 36 (which is the total number of items in the test) as the number of items attempted at 45 minutes after start. Should the two tests contain more items, a greater mean and a greater standard deviation would have been obtained. Thus the means and the standard deviations at 45 minutes after start did not reflect the actual situation and had to be discarded. From the figures for the times 15 and 30 minutes after start, it can be inferred that subjects worked on the two types of items at the same rate. However, before making such a concluding statement, one has to note that this was the first time that the subjects came across QC items. More experience with QC items may give a different result.

Chapter VI

LIMITATIONS OF THE EXPERIMENT

There are several weaknesses in this experiment and they are discussed below. It must be emphasized that the results of the experiment should be interpreted together with these weaknesses, and it is also hoped that modifications can be made to minimise or avoid these weaknesses in the event of a future study of a similar nature.

In this experiment, all the subjects came from a single school which is an Anglo-Chinese subsidised school. From past HKCEE results, it can be seen that students from this school had quite a narrow range of achievement standard (which is above average). Thus results in this experiment can only be applied to students with standard similar to the present sample. Moreover, in order that the sample would be a better representation of the population, other types of schools - Chinese Middle, government, grant-in-aid, private, technical - should also be included. Thus, in future replication, subjects of various standards drawn from different types of schools should be used.

Subjects in Group A took Test A which contained MC items that the subjects were familiar with - in their past school years, they had done thousands of this type of item. However, subjects in Group B took Test B which contained QC items. The subjects were not familiar with QC items before the experiment. Though QC items were explained, with the help of examples, to subjects in Group B before the test, their knowledge about QC items would be very limited. This inexperience with QC items would affect their performance and their rate of answering the items. Thus the results in the mean score, discrimination and the number of items attempted would be affected. A solution to this is to provide a special program for subjects taking Test B. This program should include an introduction, detailed coverage and techniques of solution of QC items. After the program, subjects would be expected to be familiar with QC items and the difference between their experience with MC and QC items would also be minimised.

The internal assessment results were used as the external criterion in the measurement of concurrent validity. The internal assessment is not a standardized test and, moreover, if subjects come from different schools, the use of a standardized test would save the trouble of comparing the standards of internal assessment in different schools. A solution is to make use of the results of the subjects in the HKCEE which take place in May every year. If the results in the HKCEE were to be used, the tests had to be administered in August, otherwise matching of subjects could not be done. However, it would be impossible to hold the tests after May when all the subjects do not have to go to school. Another solution is to administer a mathematics test and an English test (using past HKCEE papers) to the subjects. Their results in these tests would be used in the grouping and as the external criterion in the experiment.

Chapter VII

CONCLUSION AND SUGGESTIONS

The present study shows that QC items offer a reliable and valid test of mathematics achievement at the Certificate of Education level. Though some individual items may show a great deviation in the difficulty and the discrimination between MC and QC items, there is no significant difference between two tests (each containing one type of items) in these two aspects. However, it must be noted that the difficulty and the discrimination of an item depends very much on how the item is modified from the MC format. Therefore, great care must be taken in constructing QC items or modifying MC items into QC items. The rates of working for MC and QC items are the same even though subjects have never heard of QC items before the test.

On the basis of the above findings, it is found that QC items are suitable for use in the Mathematics paper in the HKCEE. However, it should be pointed out that QC items must be used together with MC items and they cannot replace MC items.

From the present study, the following question has arisen:

1. Would the use of electronic calculators affect the performance of subjects in QC items ?

2. Would it be more suitable to use QC items in testing certain mathematics topics ?

3. Would subjects of different abilities performed differently in MC and QC items ?

It is suggested that, before introducing QC items in the HKCEE, more research studies should be done in relating to the above questions. However, in future replication of the experiment, subjects taking the QC test must have sufficient practice in QC items so that their performance would not be affected by their inexperience in the items.

BIBLIOGRAPHY

Aiken, L. R. Jr. "Language factors in learning Mathematics". Review of Educational Research, Fall, 1972, Vol. 42, No. 3, 359 - 385.

Anderson, J. "Objective test in elementary analysis". Educational Studies in Mathematics, 1979, Vol. 10, 227 - 243.

Benson, J, & Crocker, L. "The effects of item format and reading ability on objective test performance: A Question on validity". Educational and Psychological Measurement, 1979, Vol. 39, 381 - 387.

Braswell, J. S. "The College Board Scholastic Aptitude Test: An overview of the mathematical portion". The Mathematics Teacher, March, 1978, Vol. 71, No.3, 168 - 180.

Cronbach, L. J. Essentials of Psychological Testing. New York, N.Y.: Harper & Row, 1970.

Ebel, R.L. Essentials of Educational Measurement. New York, N.Y.: Prentice-Hall Inc., 1979.

Ebel, R.L. "The ineffectiveness of multiple true-false items". Educational and Psychological Measurement, 1978, Vol. 38, 37 - 44.

Evans, F.R., & Pike, L.W. "The effects of instruction for three mathematics items formats". Journal of Educational Measurement, Winter, 1973, Vol. 10, No. 4, 257 - 271.

Feldt, L.S. "A test of hypothesis that Cronbach's alpha or Kuder-Richardson Coefficient Twenty is the same for two tests". Psychometrika, September, 1969, Vol, 34, No. 3, 363 - 373.

Forsyth, R.A. & Spratt, K.F. "Measuring problem-solving ability in mathematics with multiple-choice items: The effect of item format on selected item and test characteristics". Journal of Educational Measurement, Spring, 1980, Vol. 17, No.1, 31 - 43.

Frisbie, D.A. "Multiple choice versus true-false: A comparison of reliabilities and concurrent validities". Journal of Educational Measurement, Winter, 1973, Vol. 10, No.4, 297 - 304.

Guilford, J.P. & Fruchter, B. Fundamental Statistics in Psychology and Education. New York, N.Y.: McGraw-Hill, 1978.

Henrysson, Sten. "Gathering, analyzing, and using data on test items" in Thorndike, R.L. (Ed.), Educational Measurement, Washington D.C.: American Council on Education, 1971.

Hong Kong Certificate of Education Board. Hong Kong Certificate of Education Examination Sample Multiple-Choice Papers. Hong Kong: Hong Kong Certificate of Education Board, 1975.

Hong Kong Examinations Authority. Hong Kong Certificate of Education Examination Sample Multiple-Choice Papers, Vol. 3. Hong Kong: Hong Kong Examinations Authority, 1980.

Hong Kong Examinations Authority. Hong Kong Certificate of Education Examination Regulations and Syllabuses 1981. Hong Kong: Hong Kong Examinations Authority, 1979.

Lee L.M. "Ten years of multiple-choice testing in the Hong Kong Certificate of Education Examination". Hong Kong Science Teachers Journal, 1980, Vol. 8, No. 1, 39 - 47.

Lee, L.M. "Weighting of examination papers". Hong Kong Science Teachers Journal, 1979, Vol. 7, No.1, 61 - 67.

Linquist, E.F. (ed.) Educational Measurement. Washington D.C.: American Council on Education, 1966.

Ramos, R.A. & Stern, J. "Items behaviour associated with changes in the number of alternatives in multiple choice items". Journal of Educational Measurement, Winter, 1973, Vol. 10, No.4, 305 - 310.

Thorndike, R.L. & Hagen, E. Measurement and Evaluation in Psychology and Education. New York, N.Y.: John Wiley & Sons, 1956.

Tinkleman, S.N. "Planning the objective test" in Thorndike, R.L. (Ed.), Educational Measurement, Washington D.C.: American Council on Education, 1971.

Wesman, A.G. "Writing the test items" in Thorndike, R.L. (Ed.), Educational Measurement, Washington D.C.: American Council on Education, 1971.

Appendix I

Examples of Different Types of MC Items in Mathematics in Hong Kong

(a) Direct Question Type

With this type of item, the stem is in the form of a question asking for the answer to the problem presented. In a mathematics item of this type, there is one and only one option which is absolutely correct whilst the others are incorrect.

e.g. The perimeter of a sector is 16 and its angle is 2 radians. What is the area of the sector?

A. 16
B. 32
C. 64
D. 16π
E. 32π

(b) Incomplete Statement Type

This type of item is similar to the direct question type except that an incomplete statement is used. The incomplete statement together with the key will form an answering statement to the problem presented. An item of the direct question type can be easily changed into the incomplete statement type by replacing the question by an incomplete statement asking for the same answer.

e.g. The first term of an arithmetic progression is 6 and its tenth term is three times its second term. The common difference is

A. 18
B. 4
C. 3
D. 2
E. 1

(c) Multiple Completion Type

An item of this type consists of a stem followed by several responses, one or more of which may be correct. A second set of code letters indicates various possible combinations of correct responses. The examinee has to choose the set of code letters which designated the correct responses and to mark his answer accordingly. Items of this type have a characteristic that the stem does not really formulate a problem -- it only acts as a base for determining whether each of the responses that follows is correct or not. This type is useful for problems which have a number of correct answers.

e.g. The Highest Common Factor of two unequal positive integers a and b is 8. Which of the following must be true ?

(1) The difference between a and b is divisible by 8

(2) (a + b) is divisible by 16

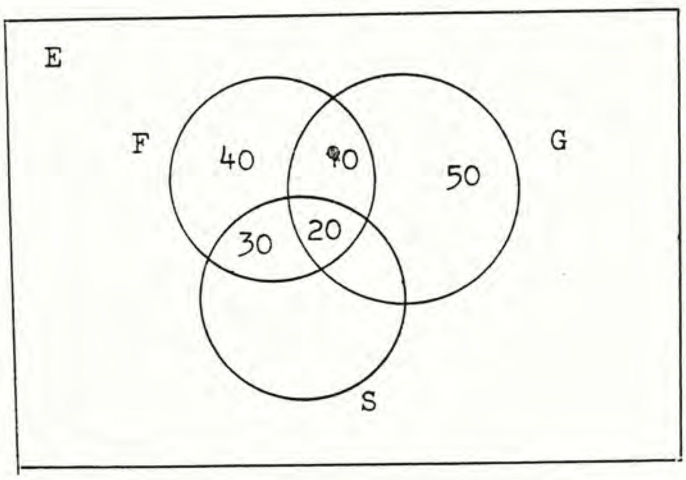
(3) ab is divisible by 64

A. (3) only
B. (1) and (2) only
C. (1) and (3) only
D. (2) and (3) only
E. (1), (2) and (3)

(d) Negative Type

With this type of item, the responses include several correct answers and one incorrect answer. Instead of choosing the correct answer, examinees are instructed by the item stem to choose the option which does not correctly answer the question. This type of item is applicable to questions that would normally have several equally good answers. Usually, the examinees' attention are drawn to the fact that an item is of the negative type by printing the word "not" in italics or bold type.

e.g.



In the Venn diagram above,
E = {students in a school}
S = {short-sighted students}
F = {form 5 students}
G = {girls}

How many form 5 girls are NOT short-sighted ?

A. 10
B. 20
C. 30
D. 40
E. 50

Appendix II

INSTRUCTIONS FOR INVIGILATORS

1. Explain the purpose of the test.
2. Emphasize that the test will not affect their internal and external examination results.
3. Give out answer sheets.
4. Give out test papers.
5. Ask students to check whether they receive the right answer sheet and test paper.
6. Ask students to read the instructions to candidates on the cover.
7. Explain to students how to put down the number of items attempted at 15 minutes, 30 minutes and 45 minutes after start.
8. Explain QC items to students taking Test B (refer to instruction and examples on front page).
9. Check the time and instruct students to start working.
10. At 15 minutes, 30 minutes and 45 minutes after start, ask students to put down the number of items attempted.
11. Five minutes before the end, reminds students.
12. Collect answer sheets and test papers separately.

Appendix III

Test A

Answer sheet for Test A

Test B

Answer sheet for Test B

Mathematics Test A

Time allowed: 1 hour

INSTRUCTIONS TO CANDIDATES

1. Answer all the questions in the order as they are presented.
2. Write down the answers in the appropriate boxes on the answer sheet.
3. At 15 minutes, 30 minutes and 45 minutes after start, fill in the number of items attempted in the appropriate boxes on the answer sheet.
4. All questions carry equal marks. No marks will be deducted for wrong answers.
5. Calculators or mathematics tables are not allowed in this test.

There are 36 questions in this test.

1. $\left(\frac{27}{64}\right)^{-\frac{2}{3}} =$

A. $\frac{4}{3}$

B. $\frac{9}{16}$

C. $\frac{16}{9}$

D. $-\frac{9}{16}$

2. If $x^2 + 5x - 6 = (x - a)(x - b)$ and $a > b$, then $a =$

A. -1
B. 1
C. 3
D. 6

3. If $2a = 3b$, then $\frac{2a^2}{3b^2} =$

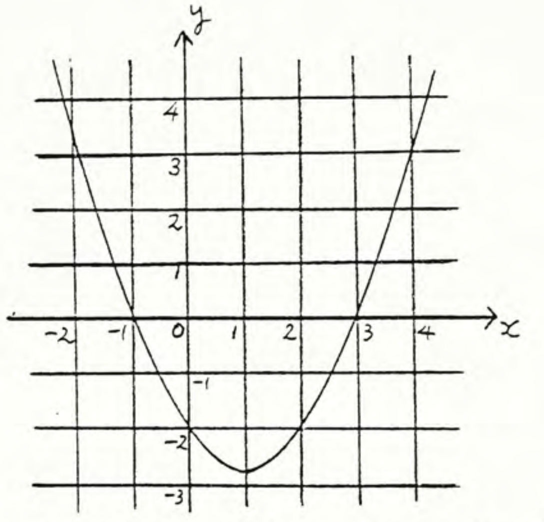
A. $\frac{9}{4}$

B. $\frac{3}{2}$

C. 1

D. $\frac{2}{3}$

4.



The figure above shows the graph of

$y = px^2 + qx + r.$

The value of r is

A. -2
B. 0
C. 2
D. 3

5. The geometric mean of x and -16 is 2. $x =$

A. 8

B. -32

C. $-\frac{1}{4}$

D. $-\frac{1}{8}$

6. When 0.001 844 81 is expressed correct to 3 significant figure, it becomes

A. 0.001 84
B. 0.001 85
C. 0.001 90
D. 0.002

7. If the height, the width and the length of a rectangular block are in the ratios of 1 : 2 : 3 respectively and its total surface area is 88 cm^2, then the height of the block is

A. 8 cm
B. 6 cm
C. 4 cm
D. 2 cm

8. If the height of a cone is halved and its base diameter is doubled, what will the new volume of the cone be ?

A. Same as the original volume
B. Half the original volume
C. Two times the original volume
D. Not sufficient data to determine

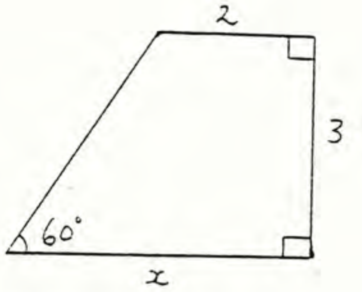
9. A man sold a car for \$35 000 at a loss of 30% on the cost price. What would have been the loss or gain percent if he had sold it for \$50 500 ?

A. A gain of 10%
B. A gain of 1%
C. No gain nor loss
D. A loss of 10 %

10. In $\triangle ABC$, $\cos A = \frac{\sqrt{3}}{2}$ and $\cos B = \frac{\sqrt{2}}{2}$. Then $\cos 2(A + B) =$

A. $\sqrt{2} + \sqrt{3}$
B. $\frac{1}{2}$
C. $\frac{\sqrt{3}}{2}$
D. $-\frac{\sqrt{3}}{2}$

11.



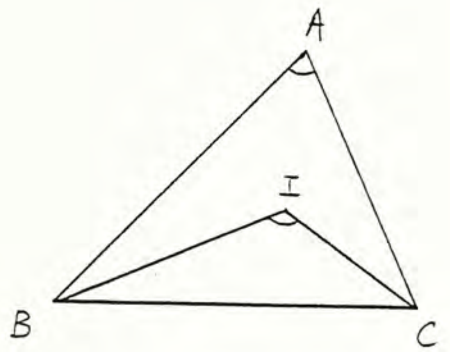
In the figure, $x =$

A. $3\frac{1}{2}$
B. $2 + 3\sqrt{3}$
C. 8
D. $2 + \sqrt{3}$

12. What is the size of the angle of a circular sector whose area is 5 cm^2 and whose radius is 10 cm ?

A. 0.05 radian
B. 0.1 radian
C. 0.5 radian
D. 2 radians

13.



In $\triangle ABC$, IB and IC are bisectors of $\angle B$ and $\angle C$ respectively. $\angle BIC$ =

A. $180° - \angle A$
B. $90° + \angle A$
C. $90° + \frac{1}{2}\angle A$
D. None of the above

14. The magnitude of $3\vec{i} - 4\vec{j}$ =

A. -1
B. 1
C. 5
D. 7

15. If $2i^2 + 3i^3 + 4i^4 + 5i^5 = a + bi$ where a and b are real numbers, then b =

A. 2
B. 8
C. -8
D. None of the above

16. In a throw of two dice, what is the probability of obtaining a total of 11 or 12 ?

A. $\frac{1}{9}$
B. $\frac{1}{12}$
C. $\frac{1}{18}$
D. $\frac{1}{36}$

17. The following are the weights in kg of 9 boys:

38 22 40 36 26 30
36 20 40

What is the median of the distribution ?

A. 26
B. 30
C. 35
D. 36

18. The slope of the straight line passing through (-3, 4) and (4, -3) is

A. -7
B. -1
C. $\frac{1}{7}$
D. 1

19. $\dfrac{3^{n+2}}{9^{n}} =$

A. 3^{2}

B. 3^{2-n}

C. 3^{n-2}

D. 3^{3n+2}

20. If $\dfrac{2x + 1}{4} = \dfrac{4 - x}{3}$,

then $x =$

A. $\dfrac{19}{10}$

B. $\dfrac{13}{10}$

C. $\dfrac{13}{7}$

D. $\dfrac{5}{7}$

21. If $x^2 + x + 1 = 4$,
then $-x^2 - x + 1 =$

A. 0
B. -2
C. -3
D. -4

22. If $25^{x} = 125$, then $x =$

A. 0.4
B. 1.5
C. 5
D. $\dfrac{2}{3}$

23. If the average of x, y and z is 4, the average of $x - 1$, $y - 5$ and $z + 3$ is

A. 1
B. 3
C. 5
D. 9

24. If $x > 0$, which of x, x^2, $\dfrac{1}{x}$ and $\sqrt{x}$ is the smallest ?

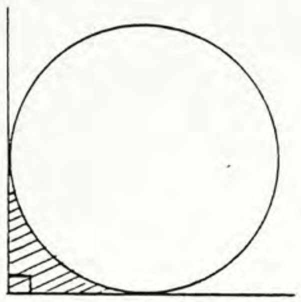
A. x

B. $\dfrac{1}{x}$

C. $\sqrt{x}$

D. Not sufficient data to determine

25.



In the figure, the radius of the circle is r. The area of the shaded part is

A. $r^2 - \frac{\pi}{4}$

B. $r(r - \frac{\pi}{4})$

C. $r^2(1 - \frac{\pi}{4})$

D. $r^2(1 - \pi)$

26. $\dfrac{\text{Volume of a sphere of radius } 2a}{\text{Volume of a sphere of radius } 3a} =$

A. $\frac{2}{3}$

B. $\frac{4}{9}$

C. $\frac{8}{27}$

D. $\frac{8\pi}{27}$

27. \$100 amounts to \$120 in 2 years at simple interest. The rate per annum is

A. 40%
B. 10%
C. 4%

D. $100\left[\left(\frac{120}{100}\right)^{\frac{1}{2}} - 1\right]\%$

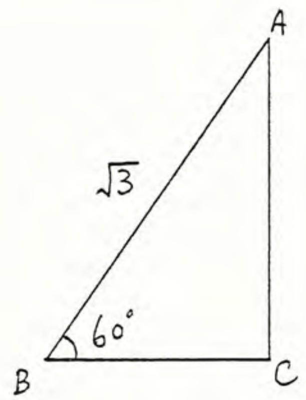
28. What is the smallest value of sin x cos y ?

A. -1

B. $-\frac{1}{2}$

C. 0
D. 1

29.

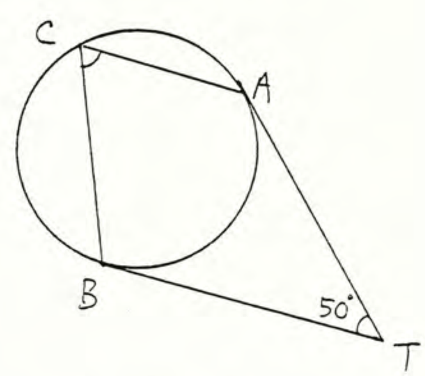


In the figure, what is the length of AC ?

A. $1\frac{1}{2}$

B. 2
C. 3
D. $\sqrt{3}$

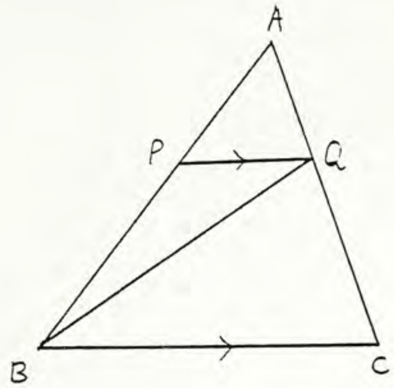
30.



In the figure, TA and TB are tangents. $\angle ATB = 50°$. $\angle ACB =$

A. 40°
B. 50°
C. 65°
D. 75°

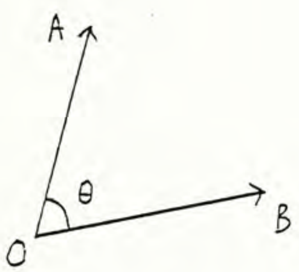
31.



In $\triangle ABC$, $PQ // BC$. The area of $\triangle APQ$ is 4. The area of $\triangle PQB$ is 6. What is the area of $\triangle QBC$?

A. 8
B. 10
C. 12
D. 15

32.

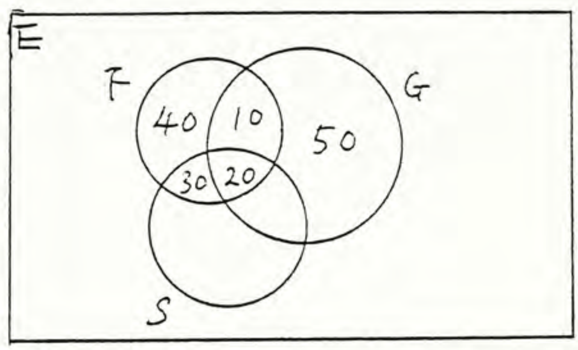


In the figure, $\overrightarrow{OA}$ and $\overrightarrow{OB}$ are each 2 units in magnitude and $\overrightarrow{OA} \cdot \overrightarrow{OB} = \frac{1}{2}$.

Then $\cos \theta$ =

A. 2
B. 1
C. $\frac{1}{8}$
D. $\frac{1}{16}$

33.



In the Venn diagram above,

E = { students in a school }
S = { short-sighted students }
F = { form 5 students }
G = { girls }

How many form 5 girls are <u>not</u> short-sighted ?

A. 10
B. 20
C. 30
D. 50

34. A bag contains 2 black balls and 2 white balls. 2 balls are taken out at random. The first ball taken out is found to be black. What is the probability that the second is white ?

A. $\frac{1}{2}$
B. $\frac{1}{3}$
C. $\frac{2}{3}$
D. $\frac{1}{4}$

35. The following table shows how Joan spends her time in a day:

sleep	9 hours
study	6 hours
recreation	5 hours
household work	1 hour
other activities	3 hours

If these data are shown in a pie chart, what is the size of the angle of the sector for recreation ?

A. $\frac{5^\circ}{24}$

B. 5°

C. 50°

D. 75°

---

36. The radius of the circle

$$x^2 + y^2 + 6x + 8y - 100 = 0$$

is

A. $5\sqrt{5}$

B. $5\sqrt{3}$

C. $10\sqrt{2}$

D. 0

END OF TEST

Mathematics Test

Answer Sheet

ame: ______________________________

chool: ______________________________

ex: Male [ ] Female [ ]

tream: Arts [ ] Science [ ]

[|]

---

or Office Use Only

(1)	(7)	(11)	(14)
[					]	1 1 1 1	0 3 6	0 0 0 0

---

rite the LETTERS of the answers in the boxes provided:

uestion No.	Answer		Question No.	Answer		Question No.	Answer		Question No.	Answer	
1	[ ]	(18)	11	[ ]	(28)	21	[ ]	(38)	31	[ ]	(48)
2	[ ]		12	[ ]		22	[ ]		32	[ ]	
3	[ ]		13	[ ]		23	[ ]		33	[ ]	
4	[ ]		14	[ ]		24	[ ]		34	[ ]	
5	[ ]		15	[ ]		25	[ ]		35	[ ]	
6	[ ]	(23)	16	[ ]	(33)	26	[ ]	(43)	36	[ ]	(53)
7	[ ]		17	[ ]		27	[ ]				
8	[ ]		18	[ ]		28	[ ]				
9	[ ]		19	[ ]		29	[ ]				
10	[ ]		20	[ ]		30	[ ]				

Time in minutes	1 5	3 0	4 5
No. of questions Attempted			

Mathematics Test B

Time allowed: 1 hour

INSTRUCTIONS TO CANDIDATES

1. Answer all the questions in the order as they are presented.
2. Write down the answers in the appropriate boxes on the answer sheet.
3. At 15 minutes, 30 minutes and 45 minutes after start, fill in the number of items attempted in the appropriate boxes on the answer sheet.
4. All questions carry equal marks. No marks will be deducted for wrong answers.
5 Calculators or mathematics tables are not allowed in this test.

QUANTITATIVE COMPARISON QUESTIONS

Directions: Each of the following questions consists of two quantities, (a) and (b).
Compare the two quantities and write on the answer sheet the letter

- A if the quantity (a) is greater;
- B if the quantity (b) is greater;
- C if the two quantities (a) and (b) are equal;
- D if the relationship cannot be determined from the information given.

Notes:

1. In certain questions, information concerning one or both of the quantities to be compared is centred above the two columns.
2. A symbol that appears in both columns represents the same thing in (a) as it does in (b).

Examples:

(i) (a) 3×6 (b) $3 + 6$ Answer: A

(ii) x, y are real numbers
(a) $x + y$ (b) $x - y$ Answer: D

There are 36 questions in this test.

1. (a) $\left(\frac{27}{64}\right)^{-\frac{2}{3}}$ (b) 1

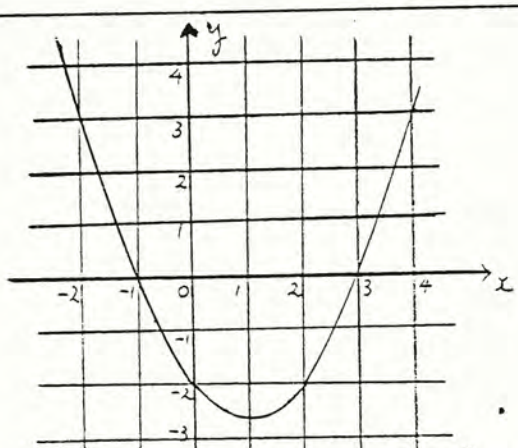
2. $x^2 + 5x - 6 = (x - a)(x - b)$ and $a > b$

(a) a (b) 6

3. $2a = 3b$

(a) $2a^2$ (b) $3b^2$

4.



The figure above shows the graph of $y = px^2 + qx + r$.

(a) r (b) -2

5. The geometric mean of x and -16 is 2.

(a) x (b) -32

6. (a) 0.001 844 81 correct to 3 significant figure (b) 0.001 84

7. The height, the width and the length of a rectangular block are in the ratio 1 : 2 : 3 respectively and its total surface area is 88 cm^2.

(a) Height of the block (b) 2 cm

8. The height of a cone is halved and its base diameter is doubled.

(a) Original volume (b) New volume

9. A man sold a car for $35 000 at a loss of 30% on the cost price.

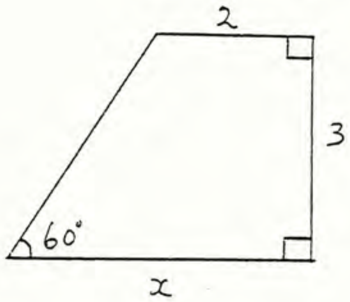
(a) His gain per cent if he sold the car for $50 500 | (b) 10%

---

10. In $\triangle ABC$, $\cos A = \frac{\sqrt{3}}{2}$ and $\cos B = \frac{\sqrt{2}}{2}$.

(a) $\cos 2(A + B)$ | (b) $-\frac{\sqrt{3}}{2}$

11.



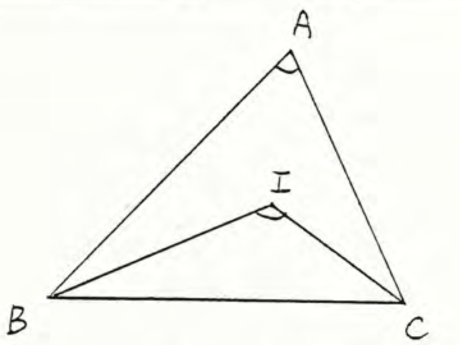
(a) x | (b) $2 + \sqrt{3}$

---

12. A circular sector has an area of 5 cm^2 and a radius of 10 cm.

(a) Angle of the sector | (b) 0.5 radian

13.



IB and IC bisects $\angle B$ and $\angle C$ respectively.

(a) $\angle BIC$ | (b) $\angle A$

---

14. (a) $|3\vec{i} - 4\vec{j}|$ | (b) $|3\vec{i} + 4\vec{j}|$

---

15. $2i^2 + 3i^3 + 4i^4 + 5i^5 = a + bi$ where a and b are real numbers.

(a) b | (b) -8

16. (a) Probability of obtaining a total of 11 or 12 in a throw of two dice (b) $\frac{1}{12}$

---

17. The weights in kg of 9 boys are as follows:

38 22 40 36 26 30 36 20 40

(a) The median of the distribution (b) 35

---

18. (a) Slope of the line joining (-3, 4) and (4, -3) (b) 1

---

19. (a) $\frac{3^{n+2}}{9^n}$ (b) 1

20.

$$\frac{2x + 1}{4} = \frac{4 - x}{3}$$

(a) x (b) $\frac{13}{7}$

21.

$$x^2 + x + 1 = 4$$

(a) $-x^2 - x + 1$ (b) -3

22.

$$25^x = 125$$

(a) x (b) 1.5

---

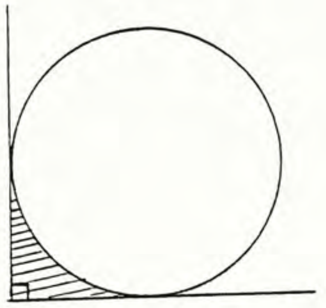
23. The average of x, y and z is 4.

(a) The average of x - 1, y - 5 and z + 3 (b) 4

24. x, x^2, $\frac{1}{x}$, $\sqrt{x}$ where $x > 0$.

(a) The smallest number (b) $\frac{1}{x}$

25.



r is the radius of the circle.

(a) Area of shaded region (b) $r^2(1 - \frac{\pi}{4})$

---

26. (a) $\dfrac{\text{Volume of a sphere of radius } 2a}{\text{Volume of a sphere of radius } 3a}$ (b) $\frac{2}{3}$

---

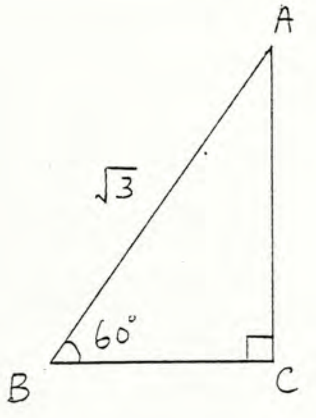
27. \$100 amount to \$120 in 2 years at simple interest.

(a) Rate per annum (b) 9.5%

---

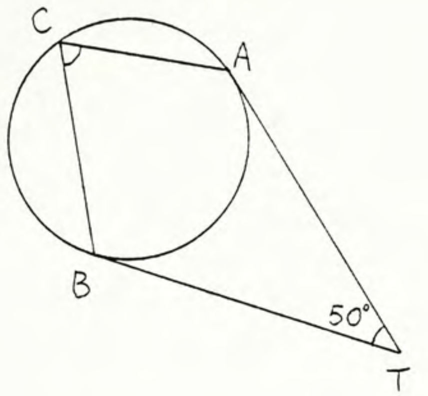
28. (a) Smallest value of $\sin x \cos y$ (b) -1

29.



(a) AC (b) 2

30.

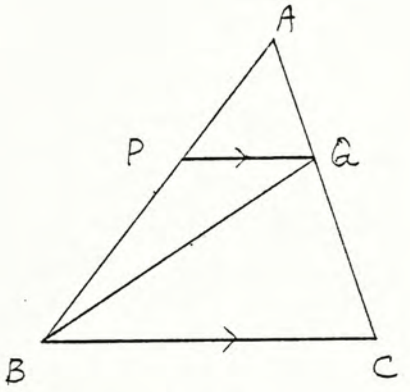


TA and TB are tangents.

(a) $\angle ACB$

(b) 60°

31.

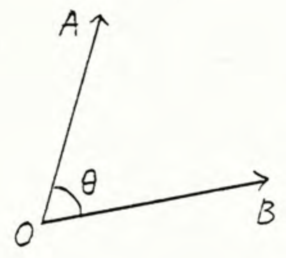


PQ//BC. The area of $\triangle APQ$ is 4 and the area of $\triangle PQB$ is 6.

(a) Area of $\triangle QBC$

(b) 15

32.

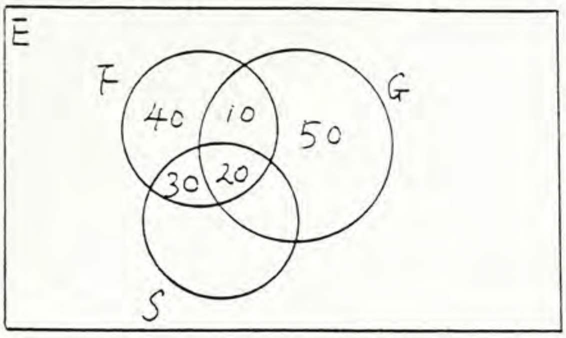


$\overrightarrow{OA}$ and $\overrightarrow{OB}$ are each 2 units in magnitude and $\overrightarrow{OA}\cdot\overrightarrow{OB} = \frac{1}{2}$.

(a) $\cos\theta$

(b) $\frac{1}{8}$

33.



In the Venn diagram,

E = { students in a school }
S = { shorted-sighted students }
F = { form 5 students }
G = { girls }

(a) Number of F.5 girls who are short-sighted

(b) Number of F.5 boys who are short-sighted.

---

34. A bag contains 2 black balls and 2 white balls. 2 balls are taken out at random. The first ball taken out is found to be black.

(a) Probability that the second ball is white

(b) $\frac{1}{2}$

---

35. The following table shows how Joan spends her time in a day:

sleep	9 hours
study	6 hours
recreation	5 hours
household work	1 hour
other activities	3 hours

The above data are represented in a pie chart.

(a) Angle of the sector for recreation

(b) 80°

---

36. (a) Radius of the circle $x^2 + y^2 + 6x + 8y - 100 = 0$

(b) $5\sqrt{3}$

END OF TEST

B

Mathematics Test

Answer Sheet

Summary of keys for answering Quantitative Comparison questions.

A	(a) is greater
B	(b) is greater
C	(a) = (b)
D	cannot be determined

Name: ______________________

School: ______________________

Sex: Male [ ] Female [ ]

Stream: Arts [ ] Science [ ]

[|]

For Office Use Only

(1)	(7)	(11)	(14)
[ ][ ][ ][ ][ ][ ]	1 1 1 1	0 3 6	0 0 0 0

Write the LETTERS of the answers in the boxes provided:

Question No.	Answer		Question No.	Answer		Question No.	Answer		Question No.	Answer	
1	[ ]	(18)	11	[ ]	(28)	21	[ ]	(38)	31	[ ]	(48)
2	[ ]		12	[ ]		22	[ ]		32	[ ]	
3	[ ]		13	[ ]		23	[ ]		33	[ ]	
4	[ ]		14	[ ]		24	[ ]		34	[ ]	
5	[ ]		15	[ ]		25	[ ]		35	[ ]	
6	[ ]	(23)	16	[ ]	(33)	26	[ ]	(43)	36	[ ]	(53)
7	[ ]		17	[ ]		27	[ ]				
8	[ ]		18	[ ]		28	[ ]				
9	[ ]		19	[ ]		29	[ ]				
10	[ ]		20	[ ]		30	[ ]				

Time in minutes	1 5	3 0	4 5
No. of questions Attempted			

Appendix IV

p-values and Coefficients of Biserial Correlation for Test A and Test B

SUBJECT/PAPER : 50 1 MATHS A

QUESTION NO.	P-VALUE	DELTA	BIS.CORR.
1	0.70	10.95	0.621
2	0.62	11.74	0.694
3	0.74	10.45	0.592
4	0.72	10.60	0.334
5	0.77	10.12	0.524
6	0.68	11.10	0.265
7	0.78	9.86	0.460
8	0.59	12.05	0.511
9	0.67	11.27	0.413
10	0.35	14.57	0.794
11	0.59	12.05	0.354
12	0.51	12.93	0.343
13	0.49	13.07	0.645
14	0.70	10.95	0.575
15	0.54	12.64	0.715
16	0.65	11.43	0.313
17	0.72	10.60	0.408
18	0.70	10.95	0.720
19	0.90	7.93	0.669
20	0.97	5.50	0.623
21	0.74	10.45	0.775
22	0.88	8.11	0.536
23	0.71	10.77	0.555
24	0.25	15.72	0.632
25	0.84	8.90	0.677
26	0.67	11.27	0.585
27	0.78	9.86	0.584
28	0.32	14.90	0.374
29	0.72	10.60	0.516
30	0.67	11.27	0.648
31	0.36	14.41	0.591
32	0.67	11.27	0.756
33	0.64	11.59	0.126
34	0.75	10.28	0.324
35	0.88	8.11	0.649
36	0.39	14.10	0.540

SUBJECT/PAPER : 60 1 MATHS B

QUESTION NO.	P-VALUE	DELTA	BIS.CORR.
1	0.62	11.74	0.561
2	0.64	11.59	0.512
3	0.80	9.73	0.244
4	0.43	13.66	0.686
5	0.83	9.18	0.511
6	0.67	11.27	0.170
7	0.33	14.73	0.487
8	0.62	11.74	0.462
9	0.54	12.64	0.458
10	0.32	14.90	0.560
11	0.39	14.10	0.577
12	0.57	12.34	0.449
13	0.83	9.18	0.450
14	0.58	12.20	0.744
15	0.67	11.27	0.453
16	0.64	11.59	0.352
17	0.58	12.20	0.398
18	0.59	12.05	0.693
19	0.67	11.27	0.317
20	0.77	10.12	0.206
21	0.54	12.64	0.672
22	0.84	8.90	0.495
23	0.80	9.73	0.569
24	0.25	15.72	0.268
25	0.67	11.27	0.741
26	0.49	13.07	0.728
27	0.75	10.28	0.715
28	0.29	15.23	0.730
29	0.86	8.37	0.386
30	0.77	10.12	0.682
31	0.20	16.27	0.782
32	0.49	13.07	0.563
33	0.73	9.36	0.428
34	0.55	12.49	0.563
35	0.91	7.45	0.616
36	0.51	12.93	0.760